



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

SWR

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/676,299	09/30/2003	Lance G. Laing	04107/100L443-US3	8447
7278	7590	07/13/2004		EXAMINER
DARBY & DARBY P.C. P. O. BOX 5257 NEW YORK, NY 10150-5257			DUNSTON, JENNIFER ANN	
			ART UNIT	PAPER NUMBER
			1636	

DATE MAILED: 07/13/2004

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)
	10/676,299	LAING, LANCE G.
Examiner	Art Unit	
Jennifer Dunston	1636	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1) Responsive to communication(s) filed on 23 February 2004.

2a) This action is **FINAL**. 2b) This action is non-final.

3) Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4) Claim(s) 1-4 is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5) Claim(s) _____ is/are allowed.

6) Claim(s) 1-4 is/are rejected.

7) Claim(s) _____ is/are objected to.

8) Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9) The specification is objected to by the Examiner.

10) The drawing(s) filed on _____ is/are: a) accepted or b) objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11) The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12) Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a) All b) Some * c) None of:

- Certified copies of the priority documents have been received.
- Certified copies of the priority documents have been received in Application No. _____.
- Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) Notice of References Cited (PTO-892)

2) Notice of Draftsperson's Patent Drawing Review (PTO-948)

3) Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)
Paper No(s)/Mail Date 11/26/2003.

4) Interview Summary (PTO-413)
Paper No(s)/Mail Date. _____.

5) Notice of Informal Patent Application (PTO-152)

6) Other: Sequence Search Attachments.

DETAILED ACTION

Claims 1-4 are pending in the instant application.

Information Disclosure Statement

Receipt of an information disclosure statement, filed on 11/23/2003, is acknowledged.

The signed and initialed PTO 1449 has been mailed with this action.

Claim Objections

Claim 4 is objected to because of the following informalities: the claim is grammatically incorrect in that the claim does not clearly indicate the numbers 4, 6, 8 and 10 are sequence identifiers. It would be remedial to amend the claim language to read something like “SEQ ID NOS: 3 and 4; SEQ ID NOS: 5 and 6.”

Claim Rejections - 35 USC § 112

The following is a quotation of the second paragraph of 35 U.S.C. 112:

The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

Claims 1-4 are rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

Claim 1 is indefinite in that the metes and bounds of the term “have” are unclear. The term “have” can be interpreted as “consisting of” or “comprising”. It would be remedial to amend the claim to use either open or closed language.

Claim 2 is indefinite in that the metes and bounds of the term “has” are unclear. The term “has” can be interpreted as “consisting of” or “comprising”. It would be remedial to amend the claim to use either open or closed language.

Claim 4 is indefinite in that the metes and bounds of the term “comprised of” are unclear. It would be remedial to amend the claim to use “consisting of”, “comprises”, or “comprising.”

Claim Rejections - 35 USC § 102

The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

Claims 1-4 are drawn to oligonucleotides comprising a sequence that differs by no more than three bases or base pairs from a sequence selected from the group consisting of SEQ ID NOS: 3-10, wherein the oligonucleotides may be double stranded and may be hybrid pairs selected from the group consisting of SEQ ID NOS: 3 and 4, SEQ ID NOS: 5 and 6, SEQ ID NOS: 7 and 8 and SEQ ID NOS: 9 and 10. The specification defines an oligonucleotide as a nucleic acid, generally of at least 10, preferably at least 15, and more preferably at least 20 nucleotides, preferably no more than 100 nucleotides, that contains a specific protein binding site (page 10, lines 6-9). Given the broadest reasonable interpretation, the claimed invention encompasses any oligonucleotide comprising a sequence selected from the group of SEQ ID NOS: 3-10.

Claims 1-4 are rejected under 35 U.S.C. 102(b) as being anticipated by Wu et al (The Journal of Biological Chemistry, Vol. 268, No. 1, pages 52-58, 1993; see the entire reference) as evidenced by San Francisco et al (Nucleic Acids Research, Vol. 18, No. 3, pages 619-624, 1990; see the entire reference).

Wu et al teach a 153 bp DNA fragment consisting of the *ars* promoter of *E. coli* plasmid R773 from nucleotides -105 to +48 (e.g. page 52, Preparation of DNA Fragments for the Gel Retardation and Footprint Assays; page 53, Promoter Region). Wu et al teach that the 153 bp DNA fragment contains the ArsR binding site (e.g. Figure 9). SEQ ID NOS: 3, 4, 7 and 8 are 100% identical to the nucleotide sequence shown within Figure 9. The sequence shown within Figure 9 is a portion of the sequence contained within plasmid pJHW1, which consists of a 0.73 kb EcoRI-HindIII fragment of pWSU1 cloned into plasmid pJBS633 (e.g. Table 1).

San Francisco et al teach plasmid pWSU1 and the 0.73 kb EcoRI-HindIII fragment of pWSU1 (e.g. page 619, Bacterial strains, plasmids and bacteriophage). San Francisco et al cloned the 0.73 kb EcoRI-HindIII fragment into plasmids pUC18 and pUC19 and phages M13mp8 and M13mp9 for sequencing (e.g. page 619 Bacterial strains, plasmids and bacteriophage). The sequence of the 0.73 kb EcoRI-HindIII fragment is contained in GenBank Accession No. X16045. The nucleic acids from bases 42-71, 71-44, 42-64 and 64-44 are 100% identical to SEQ ID NOS: 3, 4, 7 and 8 (alignments provided), respectively. Further, as indicated in the attached alignments, SEQ ID NOS: 3 and 4 and SEQ ID NOS: 7 and 8 are overlapping within GenBank Accession No. X16045.

Therefore, Wu et al necessarily teach oligonucleotides comprising SEQ ID NOS: 3, 4, 7 and 8 and hybrid pairs of SEQ ID NOS: 3 and 4 and SEQ ID NOS: 7 and 8.

Claims 1-4 are rejected under 35 U.S.C. 102(b) as being anticipated by Xu et al (The Journal of Biological Chemistry, Vol. 271, No. 5, pages 2427-2432, 1996; see the entire reference) as evidenced by Diorio et al (Journal of Bacteriology, Vol. 177, No. 8, pages 2050-2056, 1995; see the entire reference).

Xu et al teach a 208 bp fragment containing the *E. coli* chromosomal *arsR* operator (page 2429, Gel Mobility Shift and DNaseI Footprinting Assays). Xu et al generated this fragment using the forward primer 5'-CGGAATTCCGACGCAAAGTC-3' and reverse primer 5'-CCAGACGGGTTTCATCAGCAAGAATTG-3' for polymerase chain reaction amplification of the *arsR* operator region (e.g. page 2429, Primer Extension; page 2429, Gel Mobility Shift and DNaseI Footprinting Assays).

Diorio et al teach the sequence of the *E. coli* chromosomal *ars* operon (e.g. Figure 2 and GenBank Accession No. X80057). The primer sequences used by Xu et al are contained in the *E. coli* chromosomal *ars* operon sequence of GenBank Accession No. X80057. The underlined sequence of the forward primer corresponds to bases 534-546 of X80057. The reverse primer corresponds to bases 765-737 of X80057. Further, the nucleic acid sequences from bases 621-662, 662-623, 638-662 and 662-638 of X80057 are 100% identical to SEQ ID NOS: 5, 6, 9 and 10 (alignment provided), respectively. Moreover, as indicated in the attached alignments, SEQ ID NOS: 5 and 6 and SEQ ID NOS: 9 and 10 are overlapping within GenBank Accession No. X80057.

Therefore, Xu et al necessarily teach oligonucleotides comprising SEQ ID NOS: 5, 6, 9 and 10 and hybrid pairs SEQ ID NOS: 5 and 6 and SEQ ID NOS: 9 and 10.

Conclusion

No claims are allowed.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Jennifer Dunston whose telephone number is 571-272-2916. The examiner can normally be reached on M-F, 9 am to 5 pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Remy Yucel can be reached on 571-272-0781. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to (571) 272-0547.

Patent applicants with problems or questions regarding electronic images that can be viewed in the Patent Application Information Retrieval system (PAIR, <http://pair-direct.uspto.gov>) can now contact the USPTO's Patent Electronic Business Center (Patent EBC) for assistance. Representatives are available to answer your questions daily from 6 am to midnight (EST). The toll free number is (866) 217-9197. When calling please have your application serial or patent number, the type of document you are having an image problem with, the number of pages and the specific nature of the problem. The Patent Electronic Business Center will notify applicants of the resolution of the problem within 5-7 business days. Applicants can also check PAIR to confirm that the problem has been corrected. The USPTO's Patent Electronic Business Center is a complete service center supporting all patent business on the Internet. The USPTO's PAIR system provides Internet-based access to patent application status and history information. It also enables applicants to view the scanned images of their own application file folder(s) as well as general patent information available to the public.

For all other customer support, please call the USPTO Call Center (UCC) at 800-786-9199.


GERRY LEFFERS
PRIMARY EXAMINER

Jennifer Dunston
Examiner
Art Unit 1636

jad

TITLE Identification of the metalloregulatory element of the plasmid-encoded arsC resistance operon
JOURNAL Nucleic Acids Res. 18 (3), 619-624 (1990)
PMID 24080017
REFERENCE 1

AUTHORS Direct Submission
TITLE Submitted (09-AUG-1989) Rosen B.P., Department of Biochemistry, Wayne State University, School of Medicine, 54- E Canfield Avenue, Detroit MI 48201, U.S.A.

FEATURES

- source**
 - Location/Qualifiers**
 - 1. .727
 - /organism="Escherichia coli"
 - /mol_type="genomic DNA"
 - /db_xref="taxon:562"
 - /clone="PWSU1"
 - /clone_lib="pBR322"
 - 73 .79
 - /note="pot. -35 region"
 - 96 .102
 - /note="pot. -10 region"
 - 107
 - /note="transcriptional start site"
 - 114 .118
 - /note="pot. ribosome binding site"
 - 125 .478
 - /note="unnamed protein product; ArSR protein (AA 1 - 117)"
 - /codon_start=1
 - /transl_table=11
 - /protein_id="CAA34168.1"
 - /db_xref="GI:12717"
 - /db_xref="GOA:PI15905"
 - /db_xref="SWISS-PROT:PI15905"
 - /translation="MLQLPLQLPKRNLSDETRGIVLILREMGLCVCDLCMADQSQPKIRHLMRQSGILLDRKGKIVHYRLSPHPSWAQIEAWLSSQDDVAVIARKLASTYNCSSKAVCI"
 - 482 .511
 - /note="pot. stem-loop structure"
- RBS**
- Promoter**
- Promoter**
- misc feature**
- CDS**
- terminator**
- ORIGIN**

REFERENCE 2

AUTHORS Kommo,T., Kubo,A., and Nisioka,T.

TITLE Shufflon: multi-inversion of four contiguous DNA segments of plasmid R64 creates seven different open reading frames

DEFINITION AP005147

VERSION GI:20521502

KEYWORDS

SOURCE *Salmonella typhimurium*

ORGANISM *Salmonella typhimurium*

LOCUS AP005147

DEFINITION *Salmonella typhimurium* Plasmid R64 DNA, complete sequence.

VERSION AP005147

KEYWORDS

SOURCE Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; *Salmonella*.

REFERENCE 3

AUTHORS Kommo,T., Kubo,A., and Nisioka,T.

TITLE Nucleotide sequence of the rci gene encoding shufflon-specific DNA recombinase in the IncI1 plasmid R64: homology to the site-specific recombinases of the integrase family

DEFINITION AP005147

VERSION GI:20521502

KEYWORDS

SOURCE *Salmonella typhimurium*

ORGANISM *Salmonella typhimurium*

LOCUS AP005147

DEFINITION *Salmonella typhimurium* Plasmid R64 DNA, complete sequence.

VERSION AP005147

KEYWORDS

SOURCE Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; *Salmonella*.

REFERENCE 4

AUTHORS Furuya,N., Nisioka,T., and Komano,T.

TITLE Nucleotide sequence and functions of the oriT operon in IncI1

DEFINITION AP005147

VERSION GI:20521502

KEYWORDS

SOURCE *Salmonella typhimurium*

ORGANISM *Salmonella typhimurium*

LOCUS AP005147

DEFINITION *Salmonella typhimurium* Plasmid R64 DNA, complete sequence.

VERSION AP005147

KEYWORDS

SOURCE Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; *Salmonella*.

REFERENCE 5

AUTHORS Furuya,N., and Komano,T.

TITLE Determination of the nick site at oriT of IncI1 plasmid R64: global similarity of oriT structures of IncI1 and IncP plasmids

DEFINITION AP005147

VERSION GI:20521502

KEYWORDS

SOURCE *Salmonella typhimurium*

ORGANISM *Salmonella typhimurium*

LOCUS AP005147

DEFINITION *Salmonella typhimurium* Plasmid R64 DNA, complete sequence.

VERSION AP005147

KEYWORDS

SOURCE Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; *Salmonella*.

REFERENCE 6

AUTHORS Kim,S.R., Furuya,N., and Komano,T.

TITLE Nucleotide sequence and characterization of the traABCD region of IncI1 plasmid R64

DEFINITION AP005147

VERSION GI:20521502

KEYWORDS

SOURCE *Salmonella typhimurium*

ORGANISM *Salmonella typhimurium*

LOCUS AP005147

DEFINITION *Salmonella typhimurium* Plasmid R64 DNA, complete sequence.

VERSION AP005147

KEYWORDS

SOURCE Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; *Salmonella*.

REFERENCE 7

AUTHORS Furuya,N., and Komano,T.

TITLE Surface exclusion gene of IncI1 plasmid R64: nucleotide sequence and analysis of deletion mutants

DEFINITION AP005147

VERSION GI:20521502

KEYWORDS

SOURCE *Salmonella typhimurium*

ORGANISM *Salmonella typhimurium*

LOCUS AP005147

DEFINITION *Salmonella typhimurium* Plasmid R64 DNA, complete sequence.

VERSION AP005147

KEYWORDS

SOURCE Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; *Salmonella*.

REFERENCE 8

AUTHORS Furuya,N., and Komano,T.

TITLE Nucleotide sequence and characterization of the traBAC region of the IncI1 plasmid R64: global similarity of oriT structures of IncI1 and IncP plasmids

DEFINITION AP005147

VERSION GI:20521502

KEYWORDS

SOURCE *Salmonella typhimurium*

ORGANISM *Salmonella typhimurium*

LOCUS AP005147

DEFINITION *Salmonella typhimurium* Plasmid R64 DNA, complete sequence.

VERSION AP005147

KEYWORDS

SOURCE Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; *Salmonella*.

REFERENCE 9

AUTHORS Kim,S.R., and Komano,T.

TITLE The plasmid R64 thin pilus identified as a type IV pilus

DEFINITION AP005147

VERSION GI:20521502

KEYWORDS

SOURCE *Salmonella typhimurium*

ORGANISM *Salmonella typhimurium*

LOCUS AP005147

DEFINITION *Salmonella typhimurium* Plasmid R64 DNA, complete sequence.

VERSION AP005147

KEYWORDS

SOURCE Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; *Salmonella*.

REFERENCE 10

AUTHORS Narahara,K., Rahman,E., Furuya,N., and Komano,T.

TITLE Requirement of a limited segment of the sog gene for plasmid R64 conjugation

DEFINITION AP005147

VERSION GI:20521502

KEYWORDS

SOURCE *Salmonella typhimurium*

ORGANISM *Salmonella typhimurium*

LOCUS AP005147

DEFINITION *Salmonella typhimurium* Plasmid R64 DNA, complete sequence.

VERSION AP005147

KEYWORDS

SOURCE Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; *Salmonella*.

REFERENCE 11

AUTHORS Furuya,N., and Komano,T.

TITLE Mutational analysis of the R64 oriT region: requirement for precise location of the Nika-binding sequence

DEFINITION AP005147

VERSION GI:20521502

KEYWORDS

SOURCE *Salmonella typhimurium*

ORGANISM *Salmonella typhimurium*

LOCUS AP005147

DEFINITION *Salmonella typhimurium* Plasmid R64 DNA, complete sequence.

VERSION AP005147

KEYWORDS

SOURCE Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; *Salmonella*.

REFERENCE 12

AUTHORS Ogawa,T., and Komano,T.

TITLE Purification and characterization of thin pili of IncI1 plasmids ColIB-P9 and R64: formation of PilV-specific cell aggregates by type IV pilus

DEFINITION AP005147

VERSION GI:20521502

KEYWORDS

SOURCE *Salmonella typhimurium*

ORGANISM *Salmonella typhimurium*

LOCUS AP005147

DEFINITION *Salmonella typhimurium* Plasmid R64 DNA, complete sequence.

VERSION AP005147

KEYWORDS

SOURCE Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; *Salmonella*.

REFERENCE 13

AUTHORS Yoshida,T., Furuya,N., Ishikura,M., Isobe,T., Haino-Fukushima,K.,

TITLE Nucleotide sequence of the rci gene encoding shufflon-specific DNA recombinase in the IncI1 plasmid R64: homology to the site-specific recombinases of the integrase family

DEFINITION AP005147

VERSION GI:20521502

KEYWORDS

SOURCE *Salmonella typhimurium*

ORGANISM *Salmonella typhimurium*

LOCUS AP005147

DEFINITION *Salmonella typhimurium* Plasmid R64 DNA, complete sequence.

VERSION AP005147

KEYWORDS

SOURCE Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; *Salmonella*.

Copyright GenCore version 5.1.6
(c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: May 26, 2004, 15:20:44 ; Search time 389,487 Seconds
(without alignments)
3115,905 Million cell updates/sec

Title: US-10-676-299-4
Perfect score: 28
Sequence: 1 caacacataacaaaaaacatgtatt 28

Scoring table: IDENTITY_NUC
Gapop 10_0 , Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:
1: gb:ba:
2: gb_ltgt:
3: gb_in:
4: gb_cm:
5: gb_cv:
6: gb_dat:
7: gb_ph:
8: gb_pl:
9: gb_pr:
10: gb_ro:
11: gb_sts:
12: gb_sy:
13: gb_un:
14: gb_vl:
15: em_ba:
16: em_fun:
17: em_hum:
18: em_in:
19: em_mu:
20: em_om:
21: em_or:
22: em_cv:
23: em_pat:
24: em_ph:
25: em_D1:
26: em_ro:
27: em_sts:
28: em_in:
29: em_vl:
30: em_htg_hum:
31: em_htg_inv:
32: em_htg_other:
33: em_htg_mus:
34: em_htg_Dln:
35: em_htg_rod:
36: em_htg_mam:
37: em_htg_vrt:
38: em_sy:
39: em_htgo_hum:
40: em_htgo_mus:
41: em_htgo_other:
Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB ID	Description
C 1	C 28	100.0	727	1	ECRFARS	X16345 E. coli R-f
C 2	C 26.4	94.3	120826	1	AP005147	AP005147 Salmonell
C 3	C 21.6	77.1	169613	2	BX321886	BX321886 Danio rer
C 4	C 21.6	77.1	182051	2	AC007638	AC007638 Homo sapi
C 5	C 21.2	75.7	73000	2	AC020562	AC020562 Homo sapi
C 6	C 21.2	75.7	160811	2	AC020553	AC020553 Homo sapi
C 7	C 21.2	75.7	174303	10	AC144801	AC144801 Mus muscu
C 8	C 21.2	75.7	171109	2	AC118161	AC118161 Rattus no
C 9	C 21.2	75.7	184897	9	AC090919	AC090919 Homo sapi
C 10	C 21.2	75.7	199985	2	AC130151	AC130151 Rattus no
C 11	C 20.6	73.6	6391	6	AX323523	AX323523 Sequence
C 12	C 20.6	73.6	10099	1	AB002489	AB002489 Neisseria
C 13	C 20.6	73.6	65542	9	AC091766	AC091766 Homo sapi
C 14	C 20.6	73.6	116895	10	AL228957	AL228957 Mouse DNA
C 15	C 20.6	73.6	121636	2	AC122166	AC122166 Medicago
C 16	C 20.6	73.6	148157	9	AL153533	AL153533 Human DNA
C 17	C 20.6	73.6	160127	2	AC068977	AC068977 Homo sapi
C 18	C 20.6	73.6	166762	2	AC055552	AC055552 Homo sapi
C 19	C 20.6	73.6	198788	8	ATCHRIV46	AL161546 Arabidopsis
C 20	C 20.6	73.6	207674	8	ATEFCAR	Z97143 Arabidopsis
C 21	C 20.6	73.6	217584	2	AC13299	AC13299 Mus muscu
C 22	C 20.6	73.6	218580	2	AC14927	AC14927 Mus muscu
C 23	C 20.6	73.6	349880	6	AX047032	AX047032 Sequence
C 24	C 20.2	72.1	34796	3	CEP5301	CEP5301
C 25	C 20.2	72.1	1516038	9	CNS01DVT	CNS01DVT
C 26	C 20.2	72.1	176345	2	AC132913	AC132913 Mus muscu
C 27	C 20.2	72.1	180044	2	AC136679	AC136679 Mus muscu
C 28	C 20.2	72.1	181902	2	AC136286	AC136286 Homo sapi
C 29	C 20.2	72.1	185702	2	AC138246	AC138246 Mus muscu
C 30	C 20.2	72.1	224453	2	AC132833	AC132833 Rattus no
C 31	C 20.2	72.1	230082	2	AC110303	AC110303 Rattus no
C 32	C 20.2	72.1	245066	2	AC125734	AC125734 Rattus no
C 33	C 20	71.4	1553	6	AX078799	AX078799 Sequence
C 34	C 20	71.4	1553	6	BD103422	BD103422 Transcript
C 35	C 20	71.4	7276	8	AB089813	AB089813 Dauca ca
C 36	C 20	71.4	12787	1	AE015654	AE015654 Sherwanell
C 37	C 20	71.4	88557	9	AL136314	AL136314 Human DNA
C 38	C 20	71.4	105320	3	AC115680	AC115680 Dicytoste
C 39	C 20	71.4	108582	2	AC120459	AC120459 Homo sapi
C 40	C 20	71.4	119716	2	AC125480	AC125480 Medicago
C 41	C 20	71.4	131289	2	AC147778	AC147778 Pan trogl
C 42	C 20	71.4	137156	2	AC138844	AC138844 Homo sapi
C 43	C 20	71.4	138008	2	AC119008	AC119008 Rattus no
C 44	C 20	71.4	140410	2	AC139483	AC139483 Homo sapi
C 45	C 20	71.4	149030	9	AC005375	AC005375 Homo sapi

ALIGNMENTS

RESULT 1	ECRFARS/	ECRFARS/	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
					X16045	X16045.1	GI:422716 arsenical resistance; arsr gene; arsr protein; DNA-binding protein; regulatory protein; resistance gene.				
								Escherichia coli			
								Bacterium; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.			
								1 (bases 1 to 227)			
								San Francisco, M.J., Hope, C.L., Owolabi, J.B., Tisa, L.S. and Rosen, B.P.			

RESULT 2
 AP005147/C
 LOCUS AP005147
 DEFINITION 1202626 bp DNA circular BCT 14-MAY-2002
 ACCESSION AP005147
 VERSION AP005147.1
 KEYWORDS GI:20521502
 SOURCE *Salmonella* *typhimurium*
 ORGANISM *Salmonella* *typhimurium*
 Bacteria: *Proteobacteria*; *Gamma-proteobacteria*; *Enterobacteriales*; *Enterobacteriaceae*; *Salmonella*.
 1 Komano,T., Kubo,A. and Nisioka,T.
 Shufflon: multi-inversion of four contiguous DNA segments of
 plasmid R64 creates seven different open reading frames
 Nucleic Acids Res. 15 (3), 1165-1172 (1987)

REFERENCE AUTHORS TITLE
 JOURNAL J. Bacteriol.
 PUBMED 87146423
 2 Kubo,A., Kusukawa,A. and Komano,T.
 Nucleotide sequence of the *rcl* gene encoding shufflon-specific DNA
 recombinases in the IncI plasmid R64: homology to the site-specific
 recombinases of integrase family

FEATURES
 source
 /organism="Escherichia coli"
 /mol_type="genomic DNA"
 /db_xref="taxon:562"
 /clone_id="pWSU1"
 /clone_libs="pBR322"
 promoter
 73 . .79
 /note="pot. -35 region"
 96 . .102
 /note="pot. -10 region"
 107 /note="transcriptional start site"
 RBS
 114 . .118
 /note="pot. ribosome binding site"
 CDS
 125 . .128
 /note="unnamed protein product; ArSR protein (AA 1 - 117)"
 /codon_start=1
 /transl_table=11
 /protein_id="CRAS4168.1"
 /db_xref="GI:42717"
 /db_xref="GDR:PI5905"
 /note="SWISS-PROT:PI5905"
 /translation="MILQTPILQFLKNSDDETRLGIVLIREMGEILCVCDLCLMADLSQ
 PTKSRHAMILRSQGILDRKQSKWHRYLSPHIPSWAQIIEQAWLSQLQQDVQWIKRK
 LASVNCSSSKAVCI"
 482 . .511
 /note="pot. stem-loop structure"

ORIGIN
 Query Match 100.0% Score 28; DB: 1; Length 727;
 Best Local Similarity 100.0% Pred. No. 0.79;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 CAACACATACCAAAACGGATATGATT 28
 Db 71 CAACACATACCAAAACGGATATGATT 44

RESULT 2
 AP005147/C
 LOCUS AP005147
 DEFINITION *Salmonella* *typhimurium* plasmid R64 DNA, complete sequence.
 ACCESSION AP005147
 VERSION AP005147.1
 KEYWORDS GI:20521502
 SOURCE *Salmonella* *typhimurium*
 ORGANISM *Salmonella* *typhimurium*
 Bacteria: *Proteobacteria*; *Gamma-proteobacteria*; *Enterobacteriales*; *Enterobacteriaceae*; *Salmonella*.
 1 Komano,T., Kubo,A. and Nisioka,T.
 Shufflon: multi-inversion of four contiguous DNA segments of
 plasmid R64 creates seven different open reading frames
 Nucleic Acids Res. 15 (3), 1165-1172 (1987)

REFERENCE AUTHORS TITLE
 JOURNAL J. Bacteriol.
 PUBMED 3029698
 2 Kubo,A., Kusukawa,A. and Komano,T.
 Nucleotide sequence of the *rcl* gene encoding shufflon-specific DNA
 recombinases in the IncI plasmid R64: homology to the site-specific
 recombinases of integrase family

GenCore version 5.1.6
 Copyright (c) 1993 - 2004 Compugen Ltd.

score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
1	42	100.0	3492	1	ECARSRC	X80057	E. coli gene
2	42	100.0	10240	1	AE000426	AE000426	Escherich
3	42	100.0	11524	1	AE015361	AE015361	Shigella
C	4	12	100.0	179341	2	AC145934	Gallus ga
Sequence:	42	100.0	225419	1	EC0039	EC0039	E. coli chr
1 ctgcacttacatcgatcata.....tcattatgttttgactta 42			AC146183	1	Pan trogl	AE016922	Shigella
Scoring table: IDENTITY_NUC			AC005575	1	AE005575	AE005575	Escherich
Gapop 10.0 , Gapext 1.0			AE005555	1	AE005555	AE005555	Escherich
Searched: 3470272 seqs, 2167151695 residues			AE016768	1	AE016768	AE016768	Escherich
Total number of hits satisfying chosen parameters: 6940544			AF242432	10	Mus muscu	AF242432	Mus muscu
Minimum DB seq length: 0			AC090511	1	Homo sapi	AC090511	Homo sapi
Maximum DB seq length: 0			AC091915	1	Homo sapi	AC091915	Homo sapi
Post-processing: Minimum Match 0%			AC110578	1	Homo sapi	AC110578	Homo sapi
Maximum Match 100%			AC079076	2	Homo sapi	AC079076	Homo sapi
Listing first 45 summaries			AC015716	2	Homo sapi	AC015716	Homo sapi
Database : GenBmlb1:*			AC134950	2	Danio rer	AC134950	Danio rer
1: gb_ba:*			AC102260	2	Mus muscu	AC102260	Mus muscu
2: gb_htg:*			AL772194	10	Mouse DNA	AL772194	Mouse DNA
3: gb_in:*			BX321087	2	Danio rer	BX321087	Danio rer
4: gb_on:*			BX255894	2	Mus muscu	BX255894	Mus muscu
5: gb_ov:*			AC116134	2	Homo sapi	AC116134	Homo sapi
6: gb_dat:*			AC110190	2	Homo sapi	AC110190	Homo sapi
7: gb_ph:*			AC091691	2	Homo sapi	AC091691	Homo sapi
8: gb_dl:*			AL954179	5	Zebrafish	AL954179	Zebrafish
9: gb_pr:*			BV005396	2	sa09all.Y	BV005396	sa09all.Y
10: gb_to:*			AC087718	2	Homo sapi	AC087718	Homo sapi
11: gb_sts:*			AC161634	2	Mus muscu	AC161634	Mus muscu
12: gb_ty:*			AC035897	2	Human DNA	AC035897	Human DNA
13: gb_in:*			AC079316	2	Homo sapi	AC079316	Homo sapi
14: gb_vt:*			AC044915	2	Homo sapi	AC044915	Homo sapi
15: em_ba:*			AC122249	10	Mus muscu	AC122249	Mus muscu
16: em_fun:*			AC118394	2	Mus muscu	AC118394	Mus muscu
17: em_hum:*			AC098469	2	Rattus no	AC098469	Rattus no
18: em_in:*			AC110449	2	Rattus no	AC110449	Rattus no
19: em_mu:*			AC115458	2	Rattus no	AC115458	Rattus no
20: em_on:*			AC121863	2	Mus muscu	AC121863	Mus muscu
21: em_or:*			AC003297	10	Oriza sat	AC003297	Oriza sat
22: em_ov:*			AC255554	5	Zebrafish	AC255554	Zebrafish
23: em_dat:*			AC1222516	2	Mus muscu	AC1222516	Mus muscu
24: em_ph:*			AC079217	2	Mus muscu	AC079217	Mus muscu
25: em_dl:*							
26: em_ro:*							
27: em_sts:*							
28: em_un:*							
29: em_vt:*							
30: em_htg_hum:*							
31: em_htg_inv:*							
32: em_htg_other:*							
33: em_htg_mus:*							
34: em_htg_pln:*							
35: em_htg_rnd:*							
36: em_htg_nm:*							
37: em_htg_vrt:*							
38: em_htg_nm:*							
39: em_htg_mus:*							
40: em_htg_other:*							

ALIGNMENTS

RESULT 1	ECARSRC						
LOCUS							
DEFINITION							
ACCESSION							
VERSION							
KEYWORDS							
SOURCE							
ORGANISM							
ENTEROBACTERIALES							
ENTEROBACTERIACEAE							
REFERENCE							
AUTHORS							
TITLE							

Pred. No. is the number of results predicted by chance to have a

An Escherichia coli chromosomal ars operon homolog is functional in

Diorio,C.; Cai,J.; Marmor,J.; Shinder,R. and Dubow,M.S.

Copyright (c) 1993 - 2004 Compugen Ltd.

Gencore version 5.1.6
OM nucleic - nucleic search, using sw model.

Run on: May 26, 2004, 15:20:44 ; Search time 556.41 Seconds
(without alignments)

3115.905 Million cell updates/sec

Title: US-10-676-299-6

Prefect score: 40

Sequence: 1 taatccaaaaacatataatgacttaacgaaatgttaagtgc 40

Scoring table: IDENTITY_NUC

Gapop 10_0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters:

6940544 Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:*

1: gb:ba:*

2: gb:tg:*

3: gb:in:*

4: gb:cm:*

5: gb:ov:*

6: gb:pat:*

7: gb:ph:*

8: gb:pl:*

9: gb:pr:*

10: gb:ro:*

11: gb:sts:*

12: gb:sy:*

13: gb:un:*

14: gb:vi:*

15: em:ba:*

16: em:fun:*

17: em:hum:*

18: em:in:*

19: em:mu:*

20: em:om:*

21: em:or:*

22: em:ov:*

23: em:pat:*

24: em:pl:*

25: em:ro:*

26: em:sts:*

27: em:un:*

28: em:vi:*

29: em:tg:hum:*

30: em:tg:inv:*

31: em:tg:other:*

32: em:tg:mus:*

33: em:tg:pln:*

34: em:tg:rod:*

35: em:tg:nam:*

36: em:tg:vrt:*

37: em:sy:*

38: em:tg:hum:*

39: em:tg:mus:*

40: em:tg:other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	DB ID	Description	
C	40	100 0	3492	1	ECARSRC	X80057 E. coli gene	
C	40	100 0	11524	1	AE00426	AE00426 Escherichia	
C	3	40	100 0	11524	1	AE013361	AE013361 Shigella
C	4	40	100 0	179941	2	AC145934	AC145934 Gallus ga
C	5	40	100 0	235419	1	BC0UW16	U00319 E. coli chi
C	6	40	100 0	242495	2	AC146183	AC146183 Pan trogl
C	7	40	100 0	288816	1	AB016992	AB016992 Shigella
C	8	88 0	11071	1	AE005575	AE005575 Escherich	
C	9	35 2	88 0	267888	1	AP002565	AP002565 Escherich
C	10	79 0	31660	1	AB016768	AB016768 Escherich	
C	11	24 6	61 5	168112	2	AC102260	Mus muscu
C	12	24 6	61 5	176416	10	AL772194	AL772194 Mouse DNA
C	13	24 2	60 5	90650	10	AF242432	Mus muscu
C	14	24 2	60 5	156569	10	AC110190	Homo sapi
C	15	24 2	60 5	163314	9	AC091611	AC091611 Homo sapi
C	16	24 2	60 5	194366	5	AB954119	AB954119 Mus muscu
C	17	24 2	60 5	198631	2	AC116741	AC116741 Mus muscu
C	18	24	60 0	88013	9	AC095111	AC095111 Homo sapi
C	19	24	60 0	15915	9	AC091515	AC091515 Homo sapi
C	20	24	60 0	169377	2	AC110577	AC110577 Homo sapi
C	21	24	60 0	178127	2	AC079076	AC079076 Homo sapi
C	22	24	60 0	198119	2	AC015716	AC015716 Homo sapi
C	23	24	60 0	207471	10	AC122249	AC122249 Mus muscu
C	24	24	60 0	212691	2	AC118594	AC118594 Mus muscu
C	25	23 8	59 5	214795	2	AC110449	AC110449 Rattus no
C	26	23 8	59 5	233330	2	AC112548	AC112548 Rattus no
C	27	23 6	59 0	27548	5	BX323038	BX323038 Zebrafish
C	28	23 6	59 0	101904	5	AB935310	AB935310 Zebrafish
C	29	23 6	59 0	121100	8	AP005774	AP005774 Oryza sat
C	30	23 6	59 0	13070	2	AP003741	AP003741 Oryza sat
C	31	23 6	59 0	142711	10	AC121963	AC121963 Mus muscu
C	32	23 6	59 0	149128	2	AC134950	AC134950 Danio rer
C	33	23 6	59 0	15233	2	AP004297	AP004297 Oryza sat
C	34	23 6	59 0	164936	5	BX255934	BX255934 Zebrafish
C	35	23 6	59 0	168190	2	AC122516	AC122516 Mus muscu
C	36	23 6	59 0	183417	5	AB451510	AB451510 Zebrafish
C	37	23 6	59 0	184808	2	AC118619	AC118619 Mus muscu
C	38	23 6	59 0	186668	2	AC114627	AC114627 Mus muscu
C	39	23 6	59 0	196588	2	BX323037	BX323037 Danio rer
C	40	23 6	59 0	203383	2	AC133933	AC133933 Mus muscu
C	41	23 6	59 0	221146	10	AC098739	AC098739 Bacteria
C	42	23 6	59 0	224406	2	BX004858	BX004858 Danio rer
C	43	23 6	59 0	225082	10	BX548065	BX548065 Mouse DNA
C	44	23 6	59 0	228871	2	AC098604	AC098604 Rattus no
C	45	23 6	59 0	253066	2	AL935184	AL935184 Danio rer

ALIGNMENTS

RESULT 1	ECARSRC/C	LOCUS	ECARSRC	DEFINITION	E.coli genes arsR, arsB, arcS.	3492 bp
X80057	X80057	X80057	X80057	GI:510824	arsB gene; arsC gene; arsenic reductase; arsenic-inducible repressor; arsR gene.	
				VERSION		
				KEYWORDS		
				SOURCE	Escherichia coli	
				ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;	
				REFERENCE	1 Diorio, C.; Cai, J.; Marmor, J.; Shinder, R. and DuBow, M.S.	
				AUTHORS	An Escherichia coli chromosomal ars operon homolog is functional in	
				TITLE		

arsenic detoxification and is conserved in gram-negative bacteria

JOURNAL J. Bacteriol. 177 (8), 2050-2056 (1995)

MEDLINE 95238276

PUBMED 7721697

2 (bases 1 to 3492)

AUTHORS Diorio, C.

Direct Submission

JOURNAL Submitted (06-JUN-1994) C. Diorio, McGill University, 3775 University Street, Montreal, Quebec, H3A 2B4, CANADA

FEATURES Location/Qualifiers

1. .3492

/organism="Escherichia coli"

/mol_type="genomic DNA"

/strain="K12"

/db_xref="taxon:562"

repeat_region

528..538

/rpt_type=INVERTED

544..554

/rpt_type=INVERTED

655..1061

/gene="arsR"

655..660

/gene="arsR"

678..683

/gene="arsR"

697..701

/gene="arsR"

708..1061

/gene="arsR"

-10_signal

RBS

repeat_region

708..1061

/product="arsenic-inducible repressor"

/protein_id="CAAS5361.1"

/db_xref="GI:5108255"

/codon_start=1

/trans_table=11

/transl_table=11

/translation="MSFPLPIQKFLIADEPRLGIVLILLSSPLGELCVCDICTADQSQPKISRHLLALRECLLDRKQKWHRLSPHPAAKLTDEARCEQKQVQAVRN

1071..12404

/gene="arsB"

1071..1076

/gene="arsB"

1072..1084

/rpt_type=INVERTED

1087..1099

/rpt_type=INVERTED

1088..1093

repeat_region

1106..1110

/gene="arsB"

1115..2104

/gene="arsB"

-10_signal

RBS

repeat_region

1106..1110

/product="membrane-located arsenite efflux pump"

/protein_id="CAAS5362.1"

/db_xref="GI:516211"

/codon_start=1

/trans_table=11

/transl_table=11

/translation="MILIGAIFVITIVIVWQKGCGIGQATGAVLAVTGIVHGDIPVWVNTWNAATFIAVIIILLLDGSFPEWAALHVSWGNGRGRLLFTWVLLGAAVAPLFDAGDGAUILLPTVIAVLLAAGTAAFLGFIADTASPLTIVSNLYNIVSADPFGCQPERYASTMVPDIAATVPLMHLXFRDQIOPQYDMALKSPRAIKDPAFTRKQWVLLVLLLFFVBLPLGIPPSAIAAAGALIIFVVAQRGHATNGKVRGAPKQVLSLGHMLVYVGLRNAGLTXLGSVYLNVLADNGKWAATLGTGKPTPISSLATLWVLSQI

MMKPTVIGAISLDGSAASQVKEARMYANVIGCDKPKTPISSLATLWVLSQI

KNWITSWGYYFRGILMLPVLFVTLAAALRLSFTL"

gene

2417..2442

/gene="arsC"

2417..2442

/gene="arsC"

CDS

/codon_start=1

/transl_table=11

/product="arsenate reductase"

/protein_id="CAAS5363.1"

/db_xref="GI:516212"

/db_xref="GOA:P37311"

/translation="MSNTIVYHNPAGTSRNTLEMNSGTCPTIHYLETPTPDELTRLCRPSVTELELDPDKGAFSKEDKGKVDEAGRK"

repeat_region

2892..2905

/rpt_type=INVERTED

2910..2933

/rpt_type=INVERTED

ORIGIN

repeat_region

Query Match 100.0%; Score 40; DB 1; Length 3492;

Best Local Similarity 100.0%; Pred. No. 0; 0/022; Gaps 0;

Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

repeat_region

1 TAGTCAAAACATATAAGCTTAACCAATGTGTAAAGTGC 40

Db 662 TAGTCAAAACATATAAGCTTAACCAATGTGTAAAGTGC 623

RESULT 2

AE000426/c

LCUS

Escherichia coli K12 MG1655 section 316 of 400 of the complete genome.

DEFINITION AE000426

ACCESSION AE000426..1

VERSION GI:1789910

KEYWORDS Escherichia coli K12

SOURCE Escherichia coli K12

ORGANISM Escherichia coli K12

BACTERIA; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.

1 (bases 1 to 10240)

REFERENCE Blattner, F.R., Plunkett, G., III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B., and Shao, Y.

AUTHORS Blattner, F.R.

JOURNAL Science 277 (3331), 1453-1474 (1997)

PUBMED 97426517

2 (bases 1 to 10240)

REFERENCE Blattner, F.R.

AUTHORS Blattner, F.R.

JOURNAL Direct Submission

TITLE Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.

Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459

3 (bases 1 to 10240)

REFERENCE Blattner, F.R.

AUTHORS Blattner, F.R.

JOURNAL Direct Submission

TITLE Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.

Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:

4 (bases 1 to 10240)

REFERENCE Blattner, F.R.

AUTHORS Blattner, F.R.

JOURNAL Direct Submission

TITLE Submitted (13-OCT-1998) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.

COMMENT This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (E. coli Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using GenMark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 (e-mail: mark@bora.technet.edu). Open reading frames that have been correlated with genetic loci are being annotated with CG

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Om nucleic - nucleic search, using sw mode:
Run on: May 26, 2004, 15:20:44 ; Search time 319.936 Seconds
3115.905 Million cell updates/sec
Title: US-10-676-299-7
Perfect score: 23
Sequence: 1 ttaatcatatgcgttttggta 23
Scoring table: IDENTITY_NUC
Gapext 1.0 , Gapext 1.0
Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : GenEmb:
1: gb_ba: *
2: gb_htg: *
3: gb_in: *
4: gb_cm: *
5: gb_ov: *
6: gb_pat: *
7: gb_Dh: *
8: gb_Dl: *
9: gb_pr: *
10: gb_pro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_m: *
21: em_or: *
22: em_ov: *
23: em_dat: *
24: em_Dh: *
25: em_Dl: *
26: em_ro: *
27: em_sts: *
28: em_vl: *
29: em_htg_hum: *
30: em_htg_inv: *
31: em_htg_other: *
32: em_htg_nus: *
33: em_htg_pln: *
34: em_htg_rod: *
35: em_htg_nam: *
36: em_htg_vrt: *
37: em_sy: *
38: em_htgo_hum: *
39: em_htgo_mus: *
40: em_htgo_other: *
41: em_htgo_vrt: *

Pred. No. is the number of results predicted by chance to have a

Result No. Score Query Match Length DB ID Description
1 23 100.0 727 1 ECRPARSR X116045 E. coli R-f
2 23 100.0 120826 1 ECRPARSR X005147 Salmonell
3 19 82.6 176325 2 BX57181 Danio rer
4 19 82.6 184543 5 BX571681 zebrafish
5 19 82.6 215891 5 AL924692 Danio rer
6 19 82.6 219601 2 BX649641 Danio rer
7 19 82.6 253172 5 BX571883 zebrafish
8 18.8 81.7 620 AF129331 Pharetra
9 18.8 81.7 3727 AX713938 Sequence
10 18.8 81.7 3727 9 AK055877 Homo sapi
11 18.8 81.7 11447 1 AE011112 Leptospir
12 18.8 81.7 39574 3 AF314193 Drosophil
13 18.8 81.7 74831 2 AC020320 Drosophil
14 18.8 81.7 112659 9 AC010677 Homo sapi
15 18.8 81.7 119595 6 BX005186 Zebrafish
16 18.8 81.7 137635 2 BX623344 Danio rer
17 18.8 81.7 161411 AC102700 Mus muscul
18 18.8 81.7 170869 3 AC011696 Drosophil
19 18.8 81.7 171831 3 AC037473 Drosophil
20 18.8 81.7 193708 9 AC05035 Homo sapi
21 18.8 81.7 223489 5 BX248100 Zebrafish
22 18.8 81.7 228048 2 AC122883 Mus muscul
23 18.8 81.7 261690 3 AE003825 Drosophil
24 18.8 81.7 269223 6 AR408762 Sequence
25 18.8 81.7 269223 6 AC067446 Sequence
26 18.8 81.7 295094 2 BX649429 Danio rer
27 18.4 80.0 214796 3 CBRG03E09 AC044483 Caenorhab
28 18.4 80.0 34796 CBF5311 AY224116 Caenorhabdi
29 18.4 80.0 99431 8 AY224118 Medicago
30 18.4 80.0 124457 8 AC146585 Medicago
31 18.4 80.0 212669 2 AC111549 Rattus no
32 18.4 80.0 261226 2 AC094788 Rattus no
33 18.2 79.1 201 8 AJ598878 Arabidops
34 18.2 79.1 430 8 AJ566897 Arabidops
35 18.2 79.1 460 11 BX530054 Arabidops
36 18.2 79.1 466 3 BOC506884 Butthus oc
37 18.2 79.1 466 3 BOC506893 Butthus oc
38 18.2 79.1 466 3 BOC506894 Butthus oc
39 18.2 79.1 466 3 BOC506896 Butthus oc
40 18.2 79.1 466 3 BOC506897 Butthus oc
41 18.2 79.1 466 3 BOC506901 Butthus oc
42 18.2 79.1 466 3 BOC506902 Butthus oc
43 18.2 79.1 486 1 BX530044 Arabidops
44 18.2 79.1 491 3 MBU550701 Mesobuthu
45 18.2 79.1 491 3 MBU550702 Mesobuthu

ALIGNMENTS

RESULT 1

ECRPARSR Locus E. coli R-f factor R773 arsr gene. DNA linear BCT 07-SEP-1994
DEFINITION X116045 ACCESSION X116045
VERSION X116045.1 GI:42716 arsenical resistance; arsr gene; ArsR protein; DNA-binding protein;
KEYWORDS regulatory protein; resistance gene.
SOURCE Escherichia coli
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
REFERENCE 1 (bases 1 to 727)
AUTHORS San Francisco, M.J., Hope, C.L., Owolabi, J.B., Tisa, L.S. and Rosen, B.P.

FEATURES

source

1. Location/Qualifiers

1. .727

/organism="Escherichia coli"

/mol_type="genomic DNA"

/db_xref="taxon:522"

/clone="pWSU1"

/clone_lib="PBR322"

promoter

73. .79

/note="pot. -35 region"

96. .102

/note="pot. -10 region"

107

promoter

114. .118

/note="transcriptional start site"

125. .478

/note="unnamed protein product; ArsR protein (AA 1 - 117)"

CD5

1/translate_start=1

/codon_start=1

/transl_table=11

/protein_id="CAA34168.1"

/db_xref="GI:4271"

/db_xref="GOA:P15905"

/db_xref="SWISS-PROT:PI5905"

/x-translation="MLQIUTPLCFLKNTISDETRIGIVLILRENGELCVCDLGMALDQSQPKISRHLMLPESCLLDRKQGRWVHRLSPHPSWAQIIFQAWLSQQDDIVQIARKLASTNCGSGSKAVCI"

terminator

482. .511

/note="pot. stem-loop structure!"

ORIGIN

Query Match 100.0% Score 23; DB 1; Length 727;

Best Local Similarity 100.0% Pred. No. 26;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTAATATATATATCGCTTTGGTTA 23

Db 42 TTAATCATATATCGCTTTGGTTA 64

RESULT 2

AP005147 AP005147 120826 bp DNA Circular BCT 14-NAY-2002

DEFINITION *Salmonella* *typhimurium* plasmid R64 DNA, complete sequence.

ACCESSION AP005147

VERSION AP005147.1 GI:20521502

KEYWORDS

Salmonella *typhimurium*

ORGANISM *Salmonella* *typhimurium*

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; *Salmonella*.

REFERENCE 1. Komano,T., Kubo,A., and Nisioka,T.

AUTHORS Shufflon: multi-inversion of four contiguous DNA segments of plasmid R64 creates seven different open reading frames

TITLE Nucleic Acids Res. 15 (3), 1165-1172 (1987)

REFERENCE 2. Kubo,A., Kusukawa,A., and Komano,T.

AUTHORS Nucleotide sequence of the *roi* gene encoding shufflon-specific DNA recombinase in the IncI plasmid R64: homology to the site-specific recombinases of integrase family

FEATURES

source

1. Location/Qualifiers

1. .727

/organism="Escherichia coli"

/mol_type="genomic DNA"

/db_xref="taxon:522"

/clone="pWSU1"

/clone_lib="PBR322"

promoter

73. .79

/note="pot. -35 region"

96. .102

/note="pot. -10 region"

107

promoter

114. .118

/note="transcriptional start site"

125. .478

/note="unnamed protein product; ArsR protein (AA 1 - 117)"

CD5

1/translate_start=1

/codon_start=1

/transl_table=11

/protein_id="CAA34168.1"

/db_xref="GI:4271"

/db_xref="GOA:P15905"

/db_xref="SWISS-PROT:PI5905"

/x-translation="MLQIUTPLCFLKNTISDETRIGIVLILRENGELCVCDLGMALDQSQPKISRHLMLPESCLLDRKQGRWVHRLSPHPSWAQIIFQAWLSQQDDIVQIARKLASTNCGSGSKAVCI"

terminator

482. .511

/note="pot. stem-loop structure!"

ORIGIN

Query Match 100.0% Score 23; DB 1; Length 727;

Best Local Similarity 100.0% Pred. No. 26;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTAATATATATATCGCTTTGGTTA 23

Db 42 TTAATCATATATCGCTTTGGTTA 64

RESULT 2

AP005147 AP005147 120826 bp DNA Circular BCT 14-NAY-2002

DEFINITION *Salmonella* *typhimurium* plasmid R64 DNA, complete sequence.

ACCESSION AP005147

VERSION AP005147.1 GI:20521502

KEYWORDS

Salmonella *typhimurium*

ORGANISM *Salmonella* *typhimurium*

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; *Salmonella*.

REFERENCE 1. Komano,T., Kubo,A., and Nisioka,T.

AUTHORS Shufflon: multi-inversion of four contiguous DNA segments of plasmid R64 creates seven different open reading frames

TITLE Nucleic Acids Res. 15 (3), 1165-1172 (1987)

REFERENCE 2. Kubo,A., Kusukawa,A., and Komano,T.

AUTHORS Nucleotide sequence of the *roi* gene encoding shufflon-specific DNA recombinase in the IncI plasmid R64: homology to the site-specific recombinases of integrase family

FEATURES

source

1. Location/Qualifiers

1. .727

/organism="Escherichia coli"

/mol_type="genomic DNA"

/db_xref="taxon:522"

/clone="pWSU1"

/clone_lib="PBR322"

promoter

73. .79

/note="pot. -35 region"

96. .102

/note="pot. -10 region"

107

promoter

114. .118

/note="transcriptional start site"

125. .478

/note="unnamed protein product; ArsR protein (AA 1 - 117)"

CD5

1/translate_start=1

/codon_start=1

/transl_table=11

/protein_id="CAA34168.1"

/db_xref="GI:4271"

/db_xref="GOA:P15905"

/db_xref="SWISS-PROT:PI5905"

/x-translation="MLQIUTPLCFLKNTISDETRIGIVLILRENGELCVCDLGMALDQSQPKISRHLMLPESCLLDRKQGRWVHRLSPHPSWAQIIFQAWLSQQDDIVQIARKLASTNCGSGSKAVCI"

terminator

482. .511

/note="pot. stem-loop structure!"

ORIGIN

Query Match 100.0% Score 23; DB 1; Length 727;

Best Local Similarity 100.0% Pred. No. 26;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTAATATATATATCGCTTTGGTTA 23

Db 42 TTAATCATATATCGCTTTGGTTA 64

RESULT 2

AP005147 AP005147 120826 bp DNA Circular BCT 14-NAY-2002

DEFINITION *Salmonella* *typhimurium* plasmid R64 DNA, complete sequence.

ACCESSION AP005147

VERSION AP005147.1 GI:20521502

KEYWORDS

Salmonella *typhimurium*

ORGANISM *Salmonella* *typhimurium*

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; *Salmonella*.

REFERENCE 1. Komano,T., Kubo,A., and Nisioka,T.

AUTHORS Shufflon: multi-inversion of four contiguous DNA segments of plasmid R64 creates seven different open reading frames

TITLE Nucleic Acids Res. 15 (3), 1165-1172 (1987)

REFERENCE 2. Kubo,A., Kusukawa,A., and Komano,T.

AUTHORS Nucleotide sequence of the *roi* gene encoding shufflon-specific DNA recombinase in the IncI plasmid R64: homology to the site-specific recombinases of integrase family

GenCore version 5.1.6
(c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: May 26, 2004, 15:20:44 ; Search time 292.115 Seconds
(without alignments)
3115.905 Million cell updates/sec

Title: US-10-676-299-8
Perfect score: 21
Sequence: 1 taacaaaaaacgtatgatt 21

Scoring table: IDENTITY_NUC
Gapop 10_0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen Parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 10%
Listing first 45 summaries

Database : GenBmbl:
1: gb_ba:/*
2: gb_cm:/*
3: gb_in:/*
4: gb_om:/*
5: gb_or:/*
6: gb_pat:/*
7: gb_ph:/*
8: gb_P1:/*
9: gb_pr:/*
10: gb_ro:/*
11: gb_sc:/*
12: gb_sy:/*
13: gb_ur:/*
14: gb_v1:/*
15: em_ba:/*
16: em_fun:/*
17: em_hum:/*
18: em_in:/*
19: em_mu:/*
20: em_on:/*
21: em_or:/*
22: em_ov:/*
23: em_Pat:/*
24: em_ph:/*
25: em_p1:/*
26: em_ro:/*
27: em_sis:/*
28: em_ur:/*
29: em_v1:/*
30: em_hrg_hum:/*
31: em_hrg_irv:/*
32: em_hrg_other:/*
33: em_hrg_mus:/*
34: em_hrg_p1n:/*
35: em_hrg_r0d:/*
36: em_hrg_mam:/*
37: em_hrg_vrt:/*
38: em_sy:/*
39: em_hrgo_hum:/*
40: em_hrgo_mus:/*
41: em_hrgo_other:/*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	DB	ID	Description
C 1	21	100.0	727	1	ECRFARS	X16045 E. coli R-f
C 2	21	100.0	120826	1	ECRFARS	AP005147 Salmonella
C 3	18.4	87.6	21479	3	ECRFARS	AC084483 Caenorhabid
C 4	18.4	87.6	34796	3	ECRFARS	Z77956 Caenorhabidi
C 5	18.4	87.6	39574	3	ECRFARS	AF314193 Drosophil
C 6	18.4	87.6	74881	2	ECRFARS	AC020320 Drosophil
C 7	18.4	87.6	170889	3	ECRFARS	AC011696 Drosophil
C 8	18.4	87.6	171831	3	ECRFARS	AC007473 Drosophil
C 9	18.4	87.6	261690	3	ECRFARS	AB003825 Drosophil
C 10	18	85.7	96444	9	ECRFARS	AL049867 Human DNA
C 11	18	85.7	19082	2	ECRFARS	AC008057 Homo sapi
C 12	18	85.7	162494	9	ECRFARS	AC007630 Homo sapi
C 13	17.8	84.8	156	3	ECRFARS	AB506884 Birchus oc
C 14	17.8	84.8	466	3	ECRFARS	AU506893 Birchus oc
C 15	17.8	84.8	466	3	ECRFARS	AU506894 Birchus oc
C 16	17.8	84.8	466	3	ECRFARS	AU506895 Birchus oc
C 17	17.8	84.8	466	3	ECRFARS	AU506897 Birchus oc
C 18	17.8	84.8	466	3	ECRFARS	AU506901 Birchus oc
C 19	17.8	84.8	466	3	ECRFARS	AU506902 Birchus oc
C 20	17.8	84.8	491	3	ECRFARS	AU550701 Mesobuthu
C 21	17.8	84.8	491	3	ECRFARS	AU550703 Mesobuthu
C 22	17.8	84.8	491	3	ECRFARS	AU550703 Mesobuthu
C 23	17.8	84.8	599	3	ECRFARS	AY128139 Haigerda
C 24	17.8	84.8	599	3	ECRFARS	AY128139 Haigerda
C 25	17.8	84.8	603	3	ECRFARS	MTA3267 Jorunna t
C 26	17.8	84.8	658	3	ECRFARS	AZ370829 Androcton
C 27	17.8	84.8	676	3	ECRFARS	AY116607 Styela cl
C 28	17.8	84.8	778	5	ECRFARS	AY169807 Morelia v
C 29	17.8	84.8	1263	3	ECRFARS	SP4796 Cytochrome
C 30	17.8	84.8	10274	1	ECRFARS	U32797 Haemophilus
C 31	17.8	84.8	14771	3	ECRFARS	AB024528 Halocynthia
C 32	17.8	84.8	110000	2	ECRFARS	BX005451 Mus muscu
C 33	17.8	84.8	110000	6	ECRFARS	Continuation (13 o
C 34	17.8	84.8	110220	2	ECRFARS	AP003688 Orzya sat
C 35	17.8	84.8	127887	2	ECRFARS	AC125477 Medicago
C 36	17.8	84.8	131402	10	ECRFARS	AL773583 Mouse DNA
C 37	17.8	84.8	137635	2	ECRFARS	BX629344 Danio rer
C 38	17.8	84.8	162297	8	ECRFARS	AP004317 Orzya sat
C 39	17.8	84.8	165679	8	ECRFARS	AU25475 Medicago
C 40	17.8	84.8	177140	2	ECRFARS	AP005612 Orzya sat
C 41	17.8	84.8	233016	2	ECRFARS	AC113719 Rattus no
C 42	17.8	84.8	254644	2	ECRFARS	AC136836 Rattus no
C 43	17.8	84.8	256511	2	ECRFARS	AC135040 Rattus no
C 44	17.8	84.8	262721	2	ECRFARS	AC10625 Rattus no
C 45	17.8	84.8	270121	2	ECRFARS	AC132060 Rattus no

ALIGNMENTS

RESULT 1	ECRFARS/C	ECRFARS	727 bp	DNA	linear	BCT 07-SEP-1994
DEFINITION	E. coli R-factor R773 arsR Gene.					
LOCUS	X16045					
ACCESSION	X16045					
VERSION	1					
KEYWORDS	arsenal resistance; arsR gene; ArsR protein; DNA-binding protein;					
	regulatory protein; resistance gene.					
SOURCE	Escherichia coli					
ORGANISM	Escherichia coli					
	Bacteria; Proteobacteria; Escherichia; Enterobacteriales;					
REFERENCE	Enterobacteriaceae; 1 (bases 1 to 727)					
AUTHORS	San Francisco, M.J., Hope, C.L., Owolabi, J.B., Tisa, L.S. and Rosen, B.P.					

Pred. No. is the number of results predicted by chance to have a

RESULTS 2

LOCUS AP005147 C 120826 bp DNA circular BCT 14-MAY-2002
 DEFINITION Salmonella typhimurium plasmid R64 DNA, complete sequence.
 VERSION AP005147 1 GI:20521502
 KEYWORDS
 SOURCE Salmonella typhimurium
 ORGANISM Bacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Salmonellales.
 REFERENCE 1 Komano,T., Kubo,A. and Nisioka,T.
 AUTHORS Shufflin, M. multi-inversion of four contiguous DNA segments of
 Plasmid R64 creates seven different open reading frames
 Nucleic Acids Res. 15 (3), 1165-1172 (1987)
 PUBMED 87146123
 REFERENCE 2 Kubo,A., Kusukawa,A. and Komano,T.
 AUTHORS Nucleotide sequence of the rcl gene encoding shufflon-specific DNA
 recombinase in the IncI1 plasmid R64: homology to the site-specific
 recombinases of integrase family

ORIGIN

Query Match 100.0%; Score 21; DB 1; Length 727;
 Best Local Similarity 100.0%; Pred. No. 58;
 Matches 21; Conservative 0; Mismatches 0; Gaps 0;

QY	1	TAACCAAAACGCTATGATT 21
Db	64	TAACCAAAACGCTATGATT 44

RESULT 2

LOCUS AP005147 C 120826 bp DNA circular BCT 14-MAY-2002
 DEFINITION Salmonella typhimurium plasmid R64 DNA, complete sequence.
 VERSION AP005147 1 GI:20521502
 KEYWORDS
 SOURCE Salmonella typhimurium
 ORGANISM Bacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Salmonellales.
 REFERENCE 1 Komano,T., Kubo,A. and Nisioka,T.
 AUTHORS Shufflin, M. multi-inversion of four contiguous DNA segments of
 Plasmid R64 creates seven different open reading frames
 Nucleic Acids Res. 15 (3), 1165-1172 (1987)
 PUBMED 87146123
 REFERENCE 2 Kubo,A., Kusukawa,A. and Komano,T.
 AUTHORS Nucleotide sequence of the rcl gene encoding shufflon-specific DNA
 recombinase in the IncI1 plasmid R64: homology to the site-specific
 recombinases of integrase family

JOURNAL Mol. Gen. Genet. 213 (1), 30-35 (1988)
 MEDLINE 89127142
 PUBMED 3065610
 REFERENCE 3 Komano,T., Toyoshima,A., Morita,K. and Nisioka,T.
 AUTHORS Cloning and nucleotide sequence of the oriR region of the IncI1
 plasmid R64
 JOURNAL J. Bacteriol. 170 (9), 4385-4387 (1988)
 MEDLINE 88314948
 PUBMED 3045094
 REFERENCE 4 Furuya,N., Nisioka,T. and Komano,T.
 AUTHORS Nucleotide sequence and functions of the oriR operon in IncI1
 TITLE
 JOURNAL J. Bacteriol. 173 (7), 2231-2237 (1991)
 MEDLINE 9117781
 PUBMED 1848841
 REFERENCE 5 Furuya,N. and Komano,T.
 AUTHORS Determination of the nick site at oriR of IncI1 plasmid R64: global
 similarity of oriR structures of IncI1 and IncP plasmids
 JOURNAL J. Bacteriol. 173 (20), 6612-6617 (1991)
 MEDLINE 92011438
 PUBMED 1917882
 REFERENCE 6 Kim,S.R., Furuya,N. and Komano,T.
 AUTHORS Nucleotide sequence and characterization of the traABCD region of
 IncI1 plasmid R64
 JOURNAL J. Bacteriol. 175 (16), 5035-5042 (1993)
 MEDLINE 93322405
 PUBMED 8349545
 REFERENCE 7 Furuya,N. and Komano,T.
 AUTHORS Surface exclusion gene of IncI1 plasmid R64: nucleotide sequence
 and analysis of deletion mutants
 JOURNAL Plasmid 32 (1), 80-84 (1994)
 MEDLINE 95063745
 PUBMED 7991676
 REFERENCE 8 Furuya,N. and Komano,T.
 AUTHORS Nucleotide sequence and characterization of the traBABC region of
 the IncI1 plasmid R64: existence of the traB gene for plasmid
 maintenance within the transfer region
 JOURNAL J. Bacteriol. 178 (6), 1491-1497 (1996)
 MEDLINE 96198148
 PUBMED 8626273
 REFERENCE 9 Kim,S.R. and Komano,T.
 AUTHORS The plasmid R64 thin pilus identified as a type IV pilus
 JOURNAL J. Bacteriol. 179 (11), 3594-3603 (1997)
 MEDLINE 9715231
 PUBMED 9171405
 REFERENCE 10 Narahara,K., Rahman,E., Furuya,N. and Komano,T.
 AUTHORS Requirement of a limited segment of the sog gene for plasmid R64
 conjugation
 JOURNAL Plasmid 38 (1), 1-11 (1997)
 MEDLINE 9728559
 PUBMED 9281491
 REFERENCE 11 Furuya,N. and Komano,T.
 AUTHORS Mutational analysis of the Nika-binding sequence
 location of the Nika-binding sequence
 JOURNAL J. Bacteriol. 179 (23), 7291-7297 (1997)
 MEDLINE 98053841
 PUBMED 9393692
 REFERENCE 12 Yoshida,T., Furuya,N., Ishikura,M., Isobe,T., Haino-Fukushima,K.,
 Ogawa,T. and Komano,T.
 AUTHORS Purification and characterization of thin pili of IncI1 plasmids
 ColIb-P9 and R64: formation of PilV-specific cell aggregates by
 type IV Pilus
 J. Bacteriol. 180 (11), 2842-2848 (1998)
 JOURNAL

Copyright (c) 1993 - 2004 Compugen Ltd.

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

arsenic detoxification and is conserved in gram-negative bacteria
J. Bacteriol. 177 (8), 2050-2056 (1995)
9528276

PUBLMED
(bases 1 to 3492)

REFERENCE
Diorio,C.
TITLE
Direct Submission
JOURNAL
Submitted (06-JUL-1994) C. Diorio, McGill University, 3775
University Street, Montreal, Quebec, H3A 2B4, CANADA
FEATURES
source
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="K12"
/db_xref="taxon:562"

repeat_region
528. .538
/rpt_type=INVERTED

repeat_region
544. .554
/rpt_type=INVERTED

gene
655. .1061
/gene="arsR"
-35_signal
655. .660
-10_signal
678. .683
RBS
697. .701
/gene="arsR"
CDS
708. .1061
/gene="arsR"
/codon_start=1
/transl_table=11
/product="arsenic-inducible repressor"
/protein_id="CAAS6363.1"
/db_xref="GOA:P37309"
/db_xref="SWISS-PROT:P37309"
/translation="MSFLPQLPKLADBTTRIGVILLSELGELCVCDLCTALDQ
PKISRHLALIRESQSLDRLDKRQKWHYRLSPHIPAWAKKIDAEAWRCQEYQAVRN
LARQNCSDQSKNC"

gene
1071. .2444
/gene="arsB"
1071. .1076
/gene="arsB"
-35_signal
1072. .1064
repeat_region
1087. .1059
/rpt_type=INVERTED

-10_signal
1088. .1093
/gene="arsB"
RBS
1106. .1110
/gene="arsB"
CDS
1115. .2404
/gene="arsB"
/codon_start=1
/transl_table=11
/product="membrane-located arsenite efflux pump"
/protein_id="CAAS6363.1"
/db_xref="GOA:P37310"
/db_xref="SWISS-PROT:P37310"
/translation="MILAGAFVLTIVLWPKKGIGNSATIGAVLAVTGVHPG
DIPVWNNIWNNTAAFFATVIIISLILDESGEEFWALHYSRNGCRGFLFTWVILG
AAVAAFLPNDGAGLILTTAFAVMLAFLGSGTTIAFMWAGFADTASPLVTSVLY
NIVSADEFGIGREYASVMPYDIAIVATLMLYFRDIPOYDMLKSPAAI
KDPATFQIGVILLGIGLIGLIPSAIYALMELYVAKRGIAINTGSKVLR
GAFWQVTFISGMLVYGLVNGAHLGTTAEGVYVNGWAAATGIGFLTAIASSI
MNNMPTVLLGAGLSDTASGTYKEAMVYANVIGDLPKTPIGSLATLWHLVLSQ
gene
2417. .2842
/gene="arsC"
CDS
2417. .2842
/gene="arsC"
/codon_start=1

/transl_table=11
/product="arsenate reductase"
/protein_id="CAAS6363.1"
/db_xref="GOA:P37311"
/db_xref="SWISS-PROT:P37311"
/translation="MSNTIVYHNPACTSNTLLEMIRNSGTEPTIITHYLETTPPTRDEL
VKLIADMGISVRLRKNVEPEELGADEKFTDDFLIDEMLQHPLINRSTIVTSLG
TRICRPSVNLIPDAQGAFSKEDGEKVVDEAGRLK"
repeat_region
2892.
/rpt_type=INVERTED
repeat_region
2910. .2923
/rpt_type=INVERTED

ORIGIN
/db_xref="taxon:562"

repeat_region
528. .538
/rpt_type=INVERTED

repeat_region
544. .554
/rpt_type=INVERTED

gene
655. .1061
/gene="arsR"
-35_signal
655. .660
-10_signal
678. .683
RBS
697. .701
/gene="arsR"
CDS
708. .1061
/gene="arsR"
/codon_start=1
/transl_table=11
/product="arsenic-inducible repressor"
/protein_id="CAAS6363.1"
/db_xref="GOA:P37309"
/db_xref="SWISS-PROT:P37309"
/translation="MSFLPQLPKLADBTTRIGVILLSELGELCVCDLCTALDQ
PKISRHLALIRESQSLDRLDKRQKWHYRLSPHIPAWAKKIDAEAWRCQEYQAVRN
LARQNCSDQSKNC"

gene
1071. .2444
/gene="arsB"
1071. .1076
/gene="arsB"
-35_signal
1072. .1064
repeat_region
1087. .1059
/rpt_type=INVERTED

-10_signal
1088. .1093
/gene="arsB"
RBS
1106. .1110
/gene="arsB"
CDS
1115. .2404
/gene="arsB"
/codon_start=1
/transl_table=11
/product="membrane-located arsenite efflux pump"
/protein_id="CAAS6363.1"
/db_xref="GOA:P37310"
/db_xref="SWISS-PROT:P37310"
/translation="MILAGAFVLTIVLWPKKGIGNSATIGAVLAVTGVHPG
DIPVWNNIWNNTAAFFATVIIISLILDESGEEFWALHYSRNGCRGFLFTWVILG
AAVAAFLPNDGAGLILTTAFAVMLAFLGSGTTIAFMWAGFADTASPLVTSVLY
NIVSADEFGIGREYASVMPYDIAIVATLMLYFRDIPOYDMLKSPAAI
KDPATFQIGVILLGIGLIPSAIYALMELYVAKRGIAINTGSKVLR
GAFWQVTFISGMLVYGLVNGAHLGTTAEGVYVNGWAAATGIGFLTAIASSI
MNNMPTVLLGAGLSDTASGTYKEAMVYANVIGDLPKTPIGSLATLWHLVLSQ
gene
2417. .2842
/gene="arsC"
CDS
2417. .2842
/gene="arsC"
/codon_start=1

Query Match 100.0% Score 25; DB 1; Length 3492;
Best Local Similarity 100.0% Pred. No. 14;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

repeat_region
1 TTAAGTCATATGTTTGTGACTTA 25
Db 638 TTAACTATAATGTTTGTGACTTA 662

RESULT 2
AE000426 10240 bp DNA linear BCT 01-DEC-2000
DEFINITION Escherichia coli K12 MG1655 section 316 of 400 of the complete
Genome
AE000426 U00096
AE000426.1 GI:1789910

ORGANISM Escherichia coli K12
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 10240)
AUTHORS Blattner,F.R.,Plunkett,G.,III, Bloch,C.A., Perna,N.T., Burland,V.,
Riley,M.,Collado-Vides,J.,Glasner,J.D., Rode,C.K., Mayhew,G.F.,
Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
Mau,B., and Shao,Y.

TITLE The complete genome sequence of Escherichia coli K-12
JOURNAL Science 277 (5331), 1453-1474 (1997)
PUBLMED 9278503
REFERENCE 2 (bases 1 to 10240)
AUTHORS Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoli@genetics.wisc.edu Phone: 608-262-2334 Fax:
608-263-7459
REFERENCE 3 (bases 1 to 10240)
AUTHORS Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoli@genetics.wisc.edu Phone: 608-262-2334 Fax:
608-263-7459
REFERENCE 4 (bases 1 to 10240)
AUTHORS Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (13-OCT-1998) Guy Plunkett III, Laboratory of Genetics, University of
Wisconsin, 45 Henry Mall, Madison, WI 53706, USA
COMMENT This sequence was determined by the E. coli Genome Project at the
University of Wisconsin-Madison (Frederick R. Blattner, director).
Supported by NIH grants HG00301 and HG01428 (from The Human Genome
Project and NCHGR). The entire sequence was independently read by
families were determined from E. coli K12 strain MG1655. Predicted open reading
frames were determined using Genemark software, kindly supplied by
Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
30332 (e-mail: mark@member.gatech.edu). Open reading frames that
have been correlated with genetic loci are being annotated with CG

Copyright (c) 1993 - 2004 Compugen Ltd.
Gencore version 5.1.6

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	DB ID	Description
C 1	25	100.0	3492	1	ECARSRC	X80057 E. coli Gene
C 2	25	100.0	10240	1	AE000426	AE000426 Escherich
C 3	25	100.0	11524	1	AE015361	AE015361 Shigella
C 4	25	100.0	179941	2	AC145934	AC145934 Gallus ga
C 5	25	100.0	225419	1	ECOUW76	U00039 E. coli Chr
C 6	25	100.0	242495	2	AC146183	AC146183 Pan trogl
C 7	25	100.0	289816	1	AE016992	Shigella
C 8	21.8	8.2	11071	1	AE005575	AE005575 Escherich
C 9	21.8	8.2	207991	2	BX545855	BX545855 Danio rer
C 10	21.8	8.2	277888	1	AP002565	AP002565 Escherich
C 11	20.8	83.2	11661	10	AL954643	AL954643 Mouse DNA
C 12	20.8	83.2	9879	9	HS384F21	AL022171 Human DNA
C 13	20.8	83.2	112404	10	AL954640	AL954640 Mouse DNA
C 14	20.8	83.2	151950	2	AC08413	AC08413 Homo sapi
C 15	20.2	80.8	2000	6	AX508952	AX508952 Sequence
C 16	20.2	80.8	5641	6	AX346299	AX346299 Sequence
C 17	20.2	80.8	89779	8	PB005234	PB005234 Arabidops
C 18	20.2	80.8	101371	8	AC002534	AC002534 Arabidops
C 19	20.2	80.8	146710	5	AL25116	AL25116 Zebrafish
C 20	20.2	80.8	182803	2	AC14599	AC14599 Pan trogl
C 21	20.2	80.8	186981	9	AC00659	AC00659 Homo sapi
C 22	20.2	80.8	216010	2	AC119517	AC119517 Rattus no
C 23	20.2	80.8	219476	9	AC068945	AC068945 Homo sapi
C 24	20.2	80.8	220414	2	BX546182	BX546182 Danio rer
C 25	20.2	80.8	240663	2	AC097541	AC097541 Rattus no
C 26	19.8	79.2	884	11	CNS05KTW	AL43338 T3 end of
C 27	19.8	79.2	57902	1	AC103790	AC103790 Homo sapi
C 28	19.8	79.2	68181	2	AC103972	AC103972 Homo sapi
C 29	19.8	79.2	94609	8	AB046439	AB046439 Arabidops
C 30	19.8	79.2	112615	9	AC079617	AC079617 Homo sapi
C 31	19.8	79.2	115626	9	AC103792	AC103792 Homo sapi
C 32	19.8	79.2	116461	8	AC051625	AC051625 Genomic S
C 33	19.8	79.2	121038	10	AL84504	AL84504 Mouse DNA
C 34	19.8	79.2	121501	8	AC06957	AC06957 Genomic S
C 35	19.8	79.2	149098	8	AP003296	AP003296 Oryza sat
C 36	19.8	79.2	152813	2	AP05860	AP05860 Oryza sat
C 37	19.8	79.2	157295	2	AC07905	AC07905 Homo sapi
C 38	19.8	79.2	159302	2	AC098585	AC098585 Homo sapi
C 39	19.8	79.2	160295	10	AL84504	AL84504 Mouse DNA
C 40	19.8	79.2	160586	2	AC126319	AC126319 Mus muscu
C 41	19.8	79.2	176959	9	AC026402	AC026402 Homo sapi
C 42	19.8	79.2	182554	2	AC138407	AC138407 Mus muscu
C 43	19.8	79.2	183629	10	AC132596	AC132596 Mus muscu
C 44	19.8	79.2	189245	2	AC068090	AC068090 Homo sapi
C 45	19.8	79.2	195494	2	BX545912	BX545912 Danio rer

ALIGNMENTS

RESULT 1	ECARSRC/C	ECARSRC/C	LOCUS	ECARSRC
			E.coli genes arsR, arsB, arsC.	E.coli genes arsR, arsB, arsC.
			DEFINITION	
			ACCESSION	X80057
			VERSION	GI:510824
			KEYWORDS	arsB gene; arsC gene; arsenate reductase; arsenic-efflux pump;
				arsenic-inducible repressor; arsR gene.
			SOURCE	Escherichia coli
			ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
			REFERENCE	Diorio, C.; Cai, J.; Marmor, J.; Shinder, R. and DuBow, M.S.
			AUTHORS	An Escherichia coli chromosomal ars operon homolog is functional in
			TITLE	

Pred. No. is the number of results predicted by chance to have a
Pred. No. is the number of results predicted by chance to have a

